

ГРАФОВЫЕ И ТЕНЗОРНЫЕ БАЗЫ ДАННЫХ В БАНКОВСКОМ СЕКТОРЕ

КАК
УПРАВЛЯТИ
ХАОСОМ

на **20-40%**
снизить потери
от Fraud

на **15%**
Увеличить
выдачу кредитов

Заставить LLM
работать

СКРЫТЫЕ РИСКИ И УПУЩЕННЫЕ ВОЗМОЖНОСТИ: ТРИ БАРЬЕРА ДЛЯ РОСТА БАНКОВ

СКРЫТЫЕ РИСКИ В ДАННЫХ

информация о мошеннических схемах, рискованных аффилированностях, потенциальных дефолтах

- прямые финансовые потери
- репутационные риски

ЭРОЗИЯ КЛИЕНТСКОЙ ЛОЯЛЬНОСТИ

медленные, шаблонные ответы и стандартные продукты приводят к оттоку к более гибким конкурентам

- снижение Ltv и доли рынка

ПАРАЛИЧ АНАЛИТИКИ И МЕДЛЕННЫЙ CHANGE

месяцы на интеграцию данных запуска нового продукта/ проверки гипотезы. Банк не может быстро реагировать на изменения рынка

- проигрыш в конкурентной гонке



ГРАФОВЫЕ БД: ВИДИМ СУТЬ, А НЕ ТАБЛИЧКИ

ГРАФОВАЯ БАЗА ДАННЫХ

это не просто хранилище, а **цифровой двойник вашей бизнес-экосистемы**. Вместо таблиц она **оперирует сущностями и их связями**.

УЗЛЫ

Клиенты, компании, счета, транзакции, устройства, сотрудники.

СВЯЗИ

«Владеет», «Переводил деньги», «Является директором», «Использовал одно устройство».

СВОЙСТВА

	ТЕХНОЛОГИЯ	ВЛИЯНИЕ НА КРІ
Обнаружение мошенничества	Анализ транзакционных цепочек	↓ Fraud-потерь на 20-40%
Кредитный скоринг	Учет связей между клиентами	↑ Одобрение кредитов на 15%
AML	Выявление сложных схем	↓ Расходы на compliance на 30%

КЛЮЧЕВОЕ ПРЕИМУЩЕСТВО

Позволяют задавать вопросы, на которые реляционные БД отвечают слишком медленно или **не отвечают** вовсе:

- «Найди мне всех клиентов, связанных через 3+ шага с человеком из санкционного списка»
- «Кто является неформальным центром влияния в этой группе компаний?».

ТЕНЗОРНЫЕ БАЗЫ ДАННЫХ

ТЕНЗОРЫ

ЭТО **многомерные массивы данных, которые играют ключевую роль в машинном обучении (МО) и глубоком обучении.** Они позволяют эффективно хранить и обрабатывать большие объемы информации

ПРЕДСТАВЛЕНИЕ ДАННЫХ

Изображения: Представляются как 3D-тензоры.

Текст: Временные ряды или последовательности слов кодируются как 2D- или 3D-тензоры для батчей.

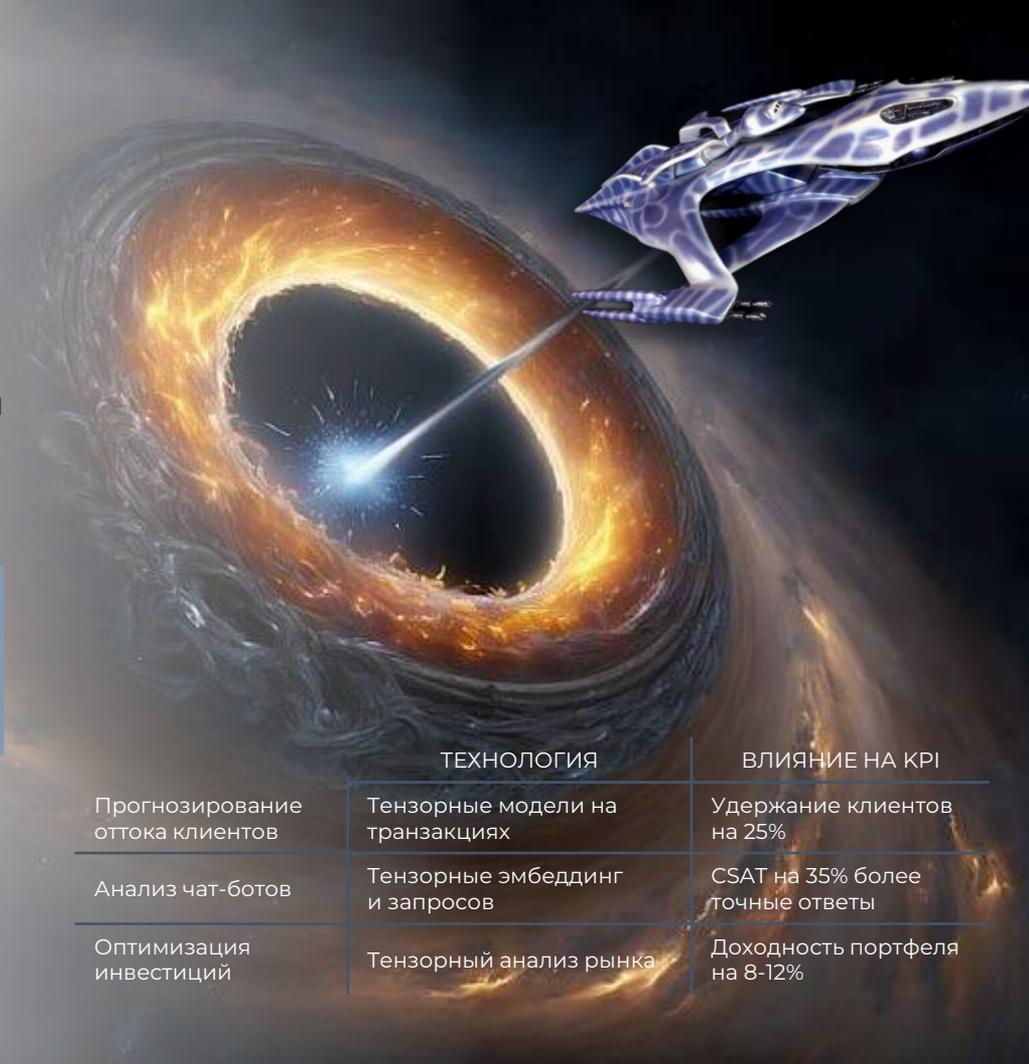
Табличные данные: 2D-тензор

ОПЕРАЦИИ В НЕЙРОННЫХ СЕТЯХ

Линейные преобразования: перемножение тензоров и матриц.

Свертки: Операции с 3D/.

Рекуррентные операции: обработка последовательностей в RNN/LSTM с использованием 3D-тензоров.



	ТЕХНОЛОГИЯ	ВЛИЯНИЕ НА КРІ
Прогнозирование оттока клиентов	Тензорные модели на транзакциях	Удержание клиентов на 25%
Анализ чат-ботов	Тензорные эмбединг и запросов	CSAT на 35% более точные ответы
Оптимизация инвестиций	Тензорный анализ рынка	Доходность портфеля на 8-12%

1 + 1 = 3: ЕДИНАЯ АРХИТЕКТУРА ДЛЯ КОНТЕКСТА И СМЫСЛА

ПРОБЛЕМА

Как **научить машину понимать смысл** текста, изображений, последовательностей действий клиента?

УЛУЧШЕНИЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ (LLM) С ПОМОЩЬЮ ГРАФОВ И ТЕНЗОРОВ:

Графы для хранения
контекстных связей

Практические
выгоды для банков

Автоматизация
сложных запросов

Тензоры для
эффективного
embedding'a и
семантического
поиска

Более точные
ответы ИИ-
ассистентов

ПОВЫШЕНИЕ КАЧЕСТВА РЕШЕНИЙ И УДОВЛЕТВОРЕННОСТИ КЛИЕНТОВ

ПРИНЯТИЕ РЕШЕНИЙ НА ВСЕХ УРОВНЯХ

- Операционный уровень: **автоматизация рутинных запросов**
- Тактический уровень: **анализ рисков** и возможностей
- Стратегический уровень: **прогнозирование трендов**

УДОВЛЕТВОРЁННОСТЬ КЛИЕНТОВ И СОТРУДНИКОВ

- **Персонализированные предложения**
- Быстрые и точные **ответы через чат-боты**
- **Уменьшение ошибок** в отчётах

КАК ЗАСТАВИТЬ LLM РАБОТАТЬ?

	ВЛИЯНИЕ НА ТОЧНОСТЬ	ВЛИЯНИЕ НА СКОРОСТЬ	МАКСИМАЛЬНЫЙ ЭФФЕКТ
Базовая LLM	60-80%	1-200 токенов/сек	-
Графы	↑ 5-15%	Нет изменений	Лучше для сложных связей
Тензоры	↑ 10-20%	Ускорение поиска	Лучше для RA
GPU-кластер	↑ 5-10%	1000+ токенов/сек	Лучше для инфер-а



ИСПОЛЬЗОВАНИЕ КОСМИЧЕСКИХ СНИМКОВ, ГРАФОВЫХ И ТЕНЗОРНЫХ БД

БАНКИ



Оценка кредитоспособности заемщиков

СТРАХОВЫЕ КОМПАНИИ



Мониторинг активов застрахованных лиц



Проверка и мониторинг залоговых объектов



Верификация и оценка страховых случаев

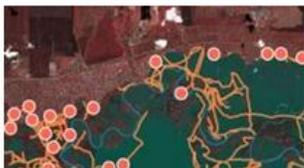
КОНСАЛТИНГ



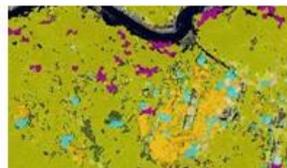
Выявление мошенничества



Моделирование и прогнозирование наступления страхового случая



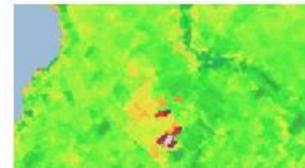
Выявление нарушений на объектах и территориях, подсчет ущерба



Контроль целевого расходования земных средств

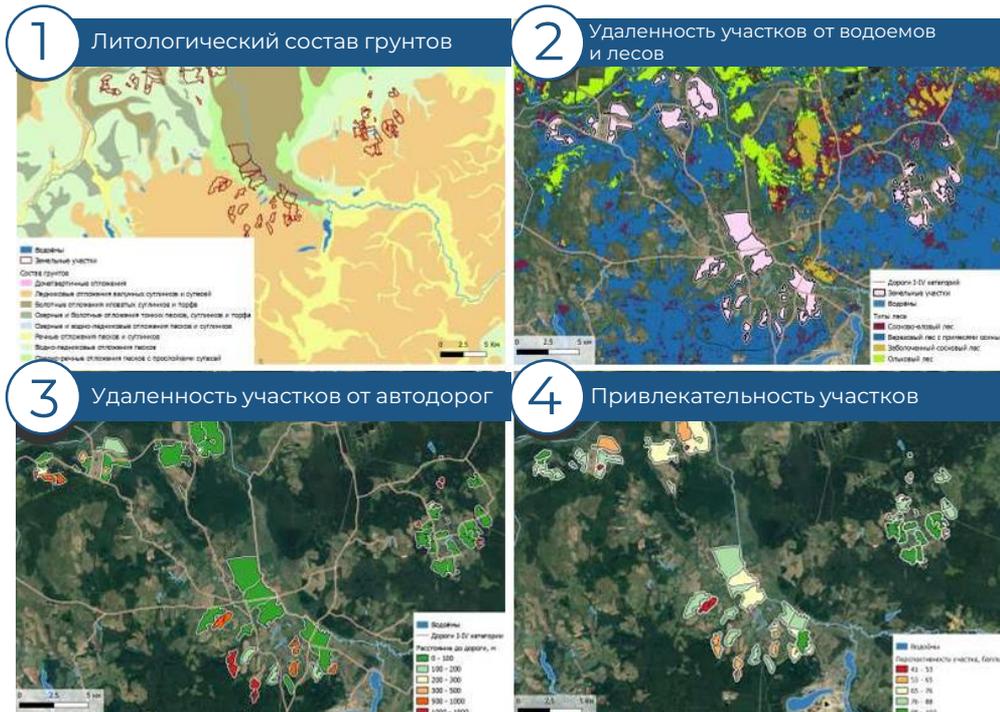


Специализированный анализ для принятия решений по финансовым операциям



Мониторинг выбросов парниковых газов

ОЦЕНКА ЗАЛОГОВОЙ НЕДВИЖИМОСТИ (АНАЛИЗ ИЗМЕНЕНИЙ ТЕРРИТОРИИ ЧЕРЕЗ СПУТНИКОВЫЕ СНИМКИ + ГРАФЫ ДЛЯ СВЯЗЕЙ ОБЪЕКТОВ)



ОЦЕНКА ЗАЛОГОВОЙ
НЕДВИЖИМОСТИ

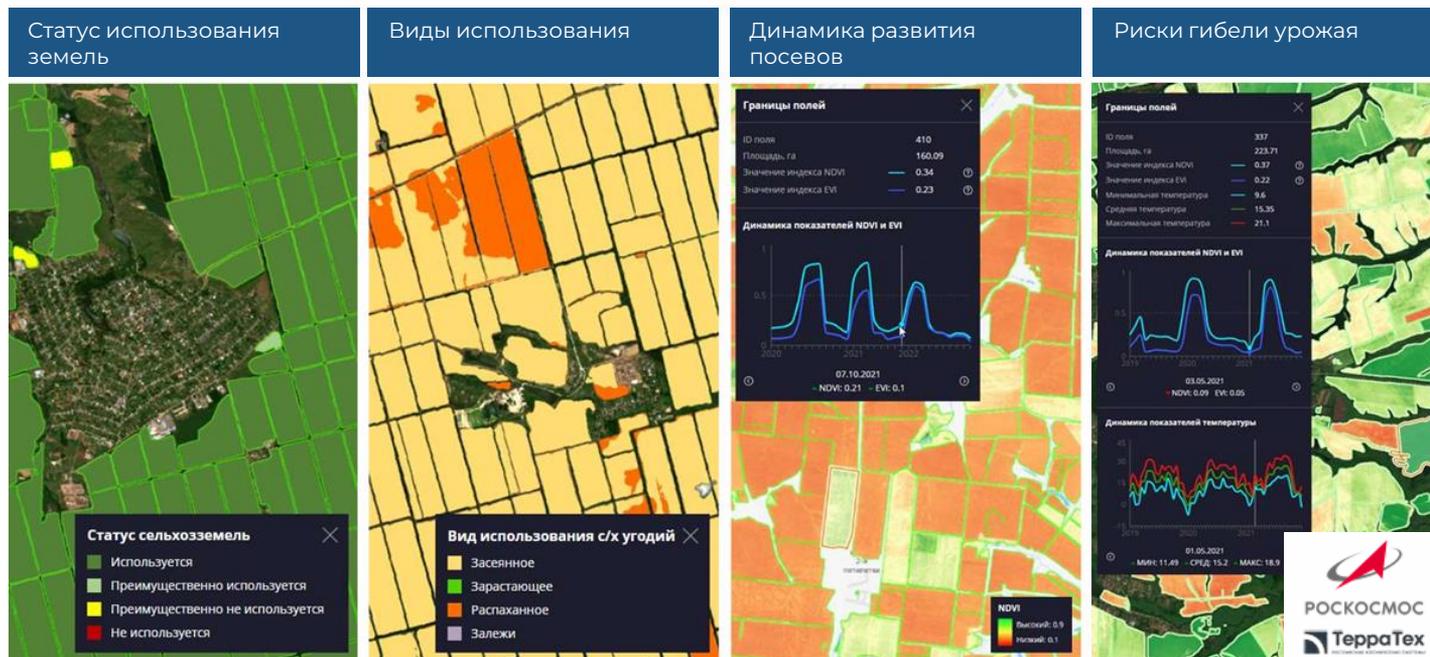
Графы (связи) + спутники (мониторинг)

↓ 50%
Просроченных залогов



МОНИТОРИНГ СЕЛЬСКОХОЗЯЙСТВЕННЫХ АКТИВОВ

(ГРАФЫ ДЛЯ УЧЕТА СВЯЗЕЙ МЕЖДУ ПОЛЯМИ, ВЛАДЕЛЬЦАМИ
И КРЕДИТАМИ)



СЕЛЬХОЗКРЕДИТОВАНИЕ

Тензоры (урожайность) + графы (схемы)

↓ 30%
NPL (плохих кредитов)

АНАЛИЗ СПУТНИКОВЫХ СНИМКОВ С ПОМОЩЬЮ ТЕНЗОРНЫХ МОДЕЛЕЙ (ВЫЯВЛЕНИЕ ИЗМЕНЕНИЙ ИНФРАСТРУКТУРЫ, ОЦЕНКА РИСКОВ НАВОДНЕНИЙ/ПОЖАРОВ)



АНАЛИЗ РИСКОВ

Система предиктивного анализа в целях обеспечения промышленной безопасности на месторождениях

Цифровая платформа, использующая геоданные, машинное обучение и аналитические модели **для прогнозирования природных и антропогенных рисков**

На основе анализа комплексных ретроспективных и оперативных данных **определяет вероятность наступления опасных процессов:**

- землетрясения
- оползневая опасность
- наводнения
- вечная мерзлота
- пожары
- удары молний
- циклоны
- терроризм
- вулканы

ПРОГНОЗИРОВАНИЕ ЭКОНОМИЧЕСКОЙ АКТИВНОСТИ

(ТЕНЗОРЫ ДЛЯ ОБРАБОТКИ ВРЕМЕННЫХ РЯДОВ СПУТНИКОВЫХ ДАННЫХ: ПОДСЧЁТ МАШИН НА ПАРКОВКАХ ТЦ, СТРОИТЕЛЬНОЙ АКТИВНОСТИ)



Готовность зданий, %

- 0% - 30%
- 60% - 85%
- 85% - 100%



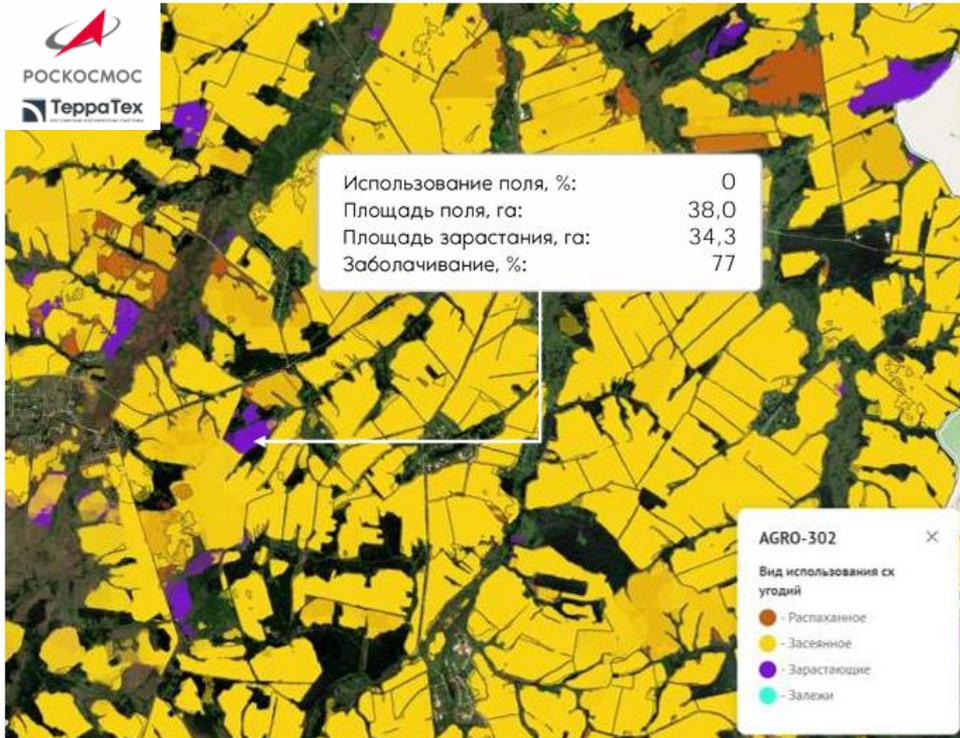
автомобильный кран



самосвал



СПУТНИКОВЫЕ ДАННЫЕ ПОКАЗЫВАЮТ СОСТОЯНИЕ ПОСЕВОВ. ТЕНЗОРНЫЕ МОДЕЛИ ПРОГНОЗИРУЮТ УРОЖАЙНОСТЬ И РИСК НЕИСПОЛНЕНИЯ ОБЯЗАТЕЛЬСТВ



КЛАССИФИКАЦИЯ ПОВЕРХНОСТИ

- Фактическое использование
- Прилегающие водные объекты
- Эрозионные процессы
- Крутизна склонов
- История затопления
- История пожаров

МОНИТОРИНГ СОСТОЯНИЯ ПОСЕВОВ

- Динамика развития растительности
- Изменение влажности поверхности
- Оперативные данные о пожарах

МЕТЕОДААННЫЕ

- Погода
- Роза ветров
- Температура

ПЕРСОНАЛИЗАЦИЯ
СЕРВИСА

ПОДКЛЮЧЕНИЕ
ПО API

КЛЮЧЕВЫЕ ВЫЗОВЫ БАНКОВ:

Высокие операционные риски → убытки

Низкая скорость принятия решений → потеря клиентов

Недостаточная персонализация → снижение NPS

Принятие решений → Финансовые результаты

Операционный уровень:

Автоматизация AML → экономия \$500K/год на ручном аудите.

Стратегический уровень:

Прогнозирование кризисов → снижение provision coverage ratio на 5%.

Удовлетворённость клиентов → Рост выручки

- NPS ↑ на 20 пунктов → рост лояльности → CLV (LTV) ↑ на 15%.

- Снижение времени обработки запросов → ↑ Конверсия в продажах на 25%.

КАК ТЕХНОЛОГИИ ПОМОГАЮТ?

Графы → снижение fraud-потерь

Тензоры → ускорение аналитики

Спутники → снижение кредитных рисков

ROI от технологий

Снижение затрат:

↓ 30%

за счёт автоматизации
compliance

↓ 20%

за счёт
снижения
fraud

Рост доходов:

↑ 15%

к cross-sell
за счёт
персонализации

↑ 10%

к прибыли
от кредитного
портфеля

ОТ ОБСУЖДЕНИЯ К ДЕЙСТВИЮ: ДАВАЙТЕ СОЗДАДИМ ДОРОЖНУЮ КАРТУ ПИЛОТОВ ДЛЯ ВАШЕГО БАНКА



Мы не предлагаем покупать "коробку".
Мы предлагаем партнерство



Проведение совместной двухдневной стратегической сессии с вашими бизнес-лидерами (B2C, B2B, Риски, Операции, Цифра)

ГОТОВЫ ЗАПУСТИТЬ ВАШ ПЕРВЫЙ CHANGE- ПРОЕКТ?





ПРИПОЖЕНИЕ

ГОД ПОСЛЕ ВНЕДРЕНИЯ: КАК ИЗМЕНИТСЯ ВАШ БАНК

к концу 1-го квартала

- Запустили пилот по AML на графах.

до **2** часов
сократилось время
расследования с 2 недель

5 млн \$
было предотвращено
потерь за квартал

на **70** %
снижено количество
ложных срабатываний,
что высвободило 5 FTE
для более сложных задач

Во 2-м квартале

- На базе графа связей мы запустили рекомендательную систему для B2B-менеджеров

На **25%**

конверсия "холодных"
звонков в целевые
встречи внутри групп
компаний выросла

+15%

к пилотномкросс-
продажам в
сегменте. (↑ Revenue)

В 3-м квартале

- Заработал умный ассистент на базе LLM и векторного поиска.

На **30%**

Нагрузка на контакт-центр
снизилась

До **85%**

показатель решения
проблемы с первого
обращения (FCR)
вырос

на **15** пунктов

Клиентская
удовлетворенность (NPS)
выросла

К КОНЦУ ГОДА

тензорная модель динамического скоринга позволила не только снизить риски, но и безопасно увеличить одобрение качественных кредитов на 10%, которые раньше получали бы отказ. (↑ Revenue, ↓ Cost of Risk)

Итоговый финансовый эффект за год рассчитывается по простой формуле и напрямую влияет на ключевой показатель для акционеров:

$$\Delta \text{EBITDA} = (\text{Пост доходов} + \text{Снижение потерь}) - (\Delta \text{OPEX} + \text{CAPEX}_{\text{год } T})$$

МЫ ПРЕДЛАГАЕМ НЕ ПРОСТО ТЕХНОЛОГИИ, А ПРОЗРАЧНУЮ И ИЗМЕРИМУЮ МОДЕЛЬ РОСТА ВАШЕЙ ПРИБЫЛЬНОСТИ

КАРТА ТРАНСФОРМАЦИИ: КАК ТЕХНОЛОГИИ ВЛИЯЮТ НА КАЖДЫЙ УРОВЕНЬ ПРИНЯТИЯ РЕШЕНИЙ

УРОВЕНЬ ПРИНЯТИЯ РЕШЕНИЙ	ПРИМЕРЫ ЗАДАЧ	ТЕХНОЛОГИЧЕСКИЙ ОТВЕТ	ВЛИЯНИЕ НА КРІ
Стратегический	<ul style="list-style-type: none">- Прогнозирование рыночных трендов- Моделирование системных рисков- Планирование M&A	<ul style="list-style-type: none">- Анализ графов новостей и рынков- Тензорные модели макроэкономики	<ul style="list-style-type: none">↑ Рост доли рынка↑ Точность прогнозов↑ Стоимость акций
Тактический	<ul style="list-style-type: none">- Оптимизация кредитного портфеля- Сегментация клиентов для кампаний- Ценообразование продуктов	<ul style="list-style-type: none">- Динамический скоринг на тензорах- Графовая кластеризация клиентов	<ul style="list-style-type: none">↑ Доходность портфеля (RoA)↑ Конверсия↑ LTV
Операционный	<ul style="list-style-type: none">- Расследование AML-инцидента- Ответ на вопрос клиента в чате- Идентификация фрода в реальном времени	<ul style="list-style-type: none">- Графовая визуализация связей- RAG-система для LLM- Анализ графа транзакций	<ul style="list-style-type: none">↓ OPEX (на compliance)↓ Стоимость контакта↓ Потери от фрода

СОЗДАЕМ ЦЕННОСТЬ ДЛЯ ЛЮДЕЙ: КЛИЕНТОВ И КОМАНДЫ



Для клиента

Переход от стандартного сервиса к "банку, который меня понимает"

Было:

"Ваш звонок очень важен для нас. Оставайтесь на линии."

→ Стало:

Проактивное предложение решения проблемы еще до обращения.

Было:

Массовая рассылка с "выгодным кредитом".

→ Стало:

Персональное предложение инвестиционного продукта, основанное на анализе целей и риск-профиля.

Результат:

↑ NPS на 20 пунктов, ↑ LTV на 15%,
↓ Отток на 25%



Для сотрудника

Переход от рутинных операций к решению творческих задач

Было:

Compliance-офицер вручную ищет данные в 5 системах.

→ Стало:

Аналитик изучает готовую визуализацию сложной мошеннической схемы и принимает решение.

Было:

Оператор поддержки ищет ответ в устаревшей базе знаний.

→ Стало:

Ассистент получает от LLM готовый, точный и персонализированный ответ для клиента.

Результат:

↑ Производительность на 40%, ↑ eNPS (удовлетворенность), ↓ Операционные ошибки

AML & ANTI-FRAUD: ОТ РЕАКТИВНОГО КОНТРОЛЯ К ПРЕДИКТИВНОЙ ЗАЩИТЕ

ТЕКУЩИЙ БИЗНЕС-ПРОЦЕСС (AS-IS):

- AML-мониторинг основан на правилах (rule-based), которые отлавливают известные паттерны.
- Расследования проводятся вручную аналитиками, которые тратят до 80% времени на сбор данных из разрозненных систем (CRM, АБС, транзакционные логи).
- Высокий уровень False Positives (до 95%), что истощает ресурсы и отвлекает от реальных угроз.

СТРАТЕГИЧЕСКИЕ ЦЕЛИ:

- Снизить финансовые потери от мошенничества и отмывания денег на 40%.
- Сократить операционные затраты на Compliance на 30%.
- Ускорить время расследования инцидентов с недель до часов.

КЛЮЧЕВЫЕ МЕТРИКИ:

- Количество продуктов на группу компаний
- Объем кросс-продаж
- Уровень просроченной задолженности (NPL) по ГК

ОГРАНИЧЕНИЯ ТЕКУЩИХ ТЕХНОЛОГИЙ (В ТОП-15 БАНКАХ РФ):

- Реляционные СУБД неэффективны для анализа связей глубже 2-3 «прыжков» (JOINS).
- Данные о клиентах, счетах, устройствах и транзакциях лежат в разных "колодцах" (data silos).
- Обработка данных в пакетном режиме (batch processing) не позволяет выявлять мошенничество в реальном времени.

КОНЦЕПЦИЯ ПИЛОТНОГО РЕШЕНИЯ (GNN-BASED ANOMALY DETECTION):

- Создание Единого Графа: данные о клиентах, транзакциях, устройствах, геолокации и сессиях загружаются в графовую БД (например, Neo4j) в режиме, близком к реальному времени.
- Поиск паттернов: алгоритмы (PageRank, Louvain modularity) выявляют аномальные кластеры, «денежных мулов» и скрытые аффилированности.
- Визуализация: аналитик получает готовую визуальную карту подозрительной активности, где подсвечены все связи, и может принять решение за минуты.

ИНТЕГРАЦИЯ И МАСШТАБИРОВАНИЕ:

- Архитектура: Пилот интегрируется с озером данных (Data Lake) или хранилищем (DWH) через CDC (Change Data Capture) коннекторы. Графовая БД становится новым «слоем обогащения данных».
- Масштабирование: После успеха пилота решение расширяется на другие виды фрода (кредитный, внутренний) и становится ядром корпоративного Anti-Fraud центра.

В2В-БАНКИНГ: УПРАВЛЕНИЕ СЛОЖНЫМИ КЛИЕНТАМИ И РОСТ КРОСС-ПРОДАЖ

ТЕКУЩИЙ БИЗНЕС-ПРОЦЕСС (AS-IS):

- Информация о группе компаний (ГК) раздроблена по разным менеджерам. Банк не видит ГК как единого клиента.
- Кредитные риски оцениваются по каждой компании в отдельности, игнорируя риски всей сети аффилированных лиц.
- Кросс-продажи основаны на интуиции менеджера, а не на данных о реальных центрах принятия решений (ЛПР).

СТРАТЕГИЧЕСКИЕ ЦЕЛИ:

- Увеличить проникновение продуктов в ключевые В2В-клиентские группы на 20%.
- Повысить точность оценки кредитного риска для ГК, снизив уровень резервирования.
- Сократить цикл продаж в В2В-сегменте.

КЛЮЧЕВЫЕ МЕТРИКИ:

- Потери от фрода
- Кол-во False Positives
- Время на расследование (Time-to-Resolution)
- Затраты на FTE в Compliance

ОГРАНИЧЕНИЯ ТЕКУЩИХ ТЕХНОЛОГИЙ (В ТОП-15 БАНКАХ РФ):

- Структуры владения и управления в CRM и АБС плоские и не отражают неформальные связи.
- Невозможно автоматически идентифицировать реальных бенефициаров и лиц, контролирурующих денежные потоки.

КОНЦЕПЦИЯ ПИЛОТНОГО РЕШЕНИЯ (GNN-BASED ANOMALY DETECTION):

- Создание графа В2В-экосистемы: В граф загружаются данные из ЕГРЮЛ, CRM, транзакций, новостных лент. Узлы — компании, люди; рёбра — «владеет», «директор», «партнёр», «денежный поток».
- Анализ влияния: Алгоритмы выявляют ключевых лиц (не всегда формальных директоров), через которых проходят финансовые потоки или которые связывают несколько компаний.
- Интерфейс для менеджера: Менеджер видит всю карту группы компаний, потенциальные риски (например, связь с банкротом) и возможности для кросс-продаж (например, предложить лизинг дочерней компании, которой он нужен).

ИНТЕГРАЦИЯ И МАСШТАБИРОВАНИЕ:

- Архитектура: Графовая БД становится центральным элементом "Customer 360" для В2В-сегмента, обогащая данные в CRM (например, Salesforce или Bitrix24).
- Масштабирование: Решение тиражируется на весь В2В-портфель, интегрируется с системами кредитного скоринга и маркетинговой автоматизации.

ДИНАМИЧЕСКИЙ СКОРИНГ И ПРЕДИКТИВНОЕ УПРАВЛЕНИЕ ОТТОКОМ

ТЕКУЩИЙ БИЗНЕС-ПРОЦЕСС (AS-IS):

- Скоринг клиентов проводится по статичным анкетным данным и кредитной истории. Обновляется редко.
- Модели оттока основаны на простых триггерах (снижение остатка на счете) и запаздывают. Банк реагирует, когда клиент уже принял решение уйти.

Стратегические цели:

- Снизить уровень клиентского оттока на 25%
- Увеличить выдачу кредитов в качественном сегменте на 15% за счет более точной оценки риска.
- Ключевые метрики: Churn Rate, Customer Lifetime Value (LTV), Approval Rate, NPL.

Ограничения текущих технологий:

- Традиционные модели машинного обучения не могут эффективно обрабатывать сложные временные последовательности (поведение клиента за год) и неструктурированные данные (тексты из чата поддержки).

Концепция пилотного решения (Behavioral Scoring & Churn Prediction):

- Создание "поведенческого вектора": С помощью тензорных вычислений (LSTM, Transformer-сети) вся последовательность транзакций, кликов в приложении, обращений в поддержку за последние 12 месяцев сворачивается в один вектор (тензор 1-го ранга) для каждого клиента.
- Предиктивная аналитика: Модель обучается на этих векторах. Она находит нелинейные паттерны, предшествующие оттоку или дефолту, за 30-60 дней до события.
- Проактивные действия: Система автоматически генерирует сигнал: "Клиент X с вероятностью 85% уйдет в следующем месяце. Рекомендуемое действие: предложить персональный вклад со ставкой Y".

Интеграция и масштабирование:

- Архитектура: Векторные представления клиентов хранятся в тензорной БД и/или в Feature Store. Модели разворачиваются на MLOps-платформе банка.
- Масштабирование: Поведенческие тензоры становятся основой для всех персонализированных коммуникаций: рекомендаций, маркетинговых кампаний, динамического ценообразования.

УМНЫЙ ПОИСК И LLM: СОЗДАНИЕ «КОРПОРАТИВНОГО МОЗГА» БАНКА

ТЕКУЩИЙ БИЗНЕС-ПРОЦЕСС (AS-IS):

- Внутренняя база знаний (для сотрудников) и FAQ (для клиентов) работают по ключевым словам. Найти нужный документ
- Чат-боты и LLM (Large Language Models) отвечают шаблонно, не знают контекста клиента и часто "галлюцинируют", выдумывая факты.

ТЕКУЩИЙ БИЗНЕС-ПРОЦЕСС (AS-IS):

- Повысить точность ответов LLM-ассистентов с 68% до 90%+
- Снизить нагрузку на 1-ю и 2-ю линии поддержки на 40%
- Ускорить онбординг новых сотрудников за счет мгновенного доступа к информации.
- Ключевые метрики: First Contact Resolution (FCR), Cost per interaction, Employee Satisfaction (eNPS), LLM Answer Accuracy.

ОГРАНИЧЕНИЯ ТЕКУЩИХ ТЕХНОЛОГИЙ

- LLM не имеют доступа к актуальной внутренней информации банка (регламенты, тарифы, история клиента).
- Поиск по ключевым словам не понимает семантику запроса ("условия досрочного погашения" и "как раньше закрыть кредит" — для него разные вещи).

Концепция пилотного решения (RAG - Retrieval-Augmented Generation):

Векторизация знаний: Вся внутренняя база знаний (регламенты, описания продуктов, скрипты) индексируется и превращается в векторы, которые загружаются в векторную БД.

Интеграция и масштабирование:

Архитектура: Графовые и тензорные БД становятся ядром корпоративной поисковой системы. Решение интегрируется с чат-платформами, CRM и внутренним порталом.

Двухэтапный ответ LLM:

Шаг 1 (Retrieval): Когда клиент или сотрудник задает вопрос, система сначала ищет в векторной БД наиболее релевантные фрагменты документов (по смыслу, а не по словам).

Масштабирование: Технология применяется для анализа звонков (Speech-to-Text -> Graph Search), автоматической классификации клиентских обращений и создания саммари по встречам.

Шаг 2 (Generation): Найденные фрагменты передаются LLM вместе с первоначальным вопросом в качестве контекста. LLM генерирует ответ, основываясь на этих точных данных, а не на своих "знаниях" из интернета.